Genomic resources for Arabidopsis research at NCBI

National Center for Biotechnology Information . National Library of Medicine . National Institutes of Health

Anjana R. Vatsan, Terence D. Murphy, Valerie Schneider, Brian Smith-White, Tatiana Tatusova, Kim D. Pruitt info@ncbi.nlm.nih.gov

ABSTRACT

aerial_tissues_BS_seq

Gene ID: 838028, updated on 7-Dec-201-

NHX8 sodium/hydrogen exchanger 8 [Arabidopsis thaliana (thale cress)]

AtNHX8, a member of the monovalent cation: proton antiporter-1 family in Arabic

GeneRIFs: Gene References Into Functions What's a GeneRIF?

An R, et al. Plant J, 2007 Feb. PMID 17270011

General protein information

NCBI Reference Sequences (RefSeq)

Reference assemble

NCBI's Entrez system is an integrated search and retrieval system that provides access to a diverse set of 39 molecular and literature databases. Arabidopsis thaliana data is available in several NCBI resources ranging from primary data submissions, meta-data resources, and core resources that provide organized views of genomes, genes, maps, and sequences. Data gets submitted to one of NCBI's archival databases - GenBank, Assembly, BioProject, BioSample, GEO, SRA, TSA, dbSNP, PopSet, Epigenomics - and then through a combination of manual curation and computational methods is integrated into NCBI's secondary databases like, BLAST, CloneDB, Gene, RefSeq, Map Viewer, Genome, HomoloGene, UniGene, and Protein Clusters. Over 2100 Arabidopsis BioProjects are registered with NCBI; these projects reflect the areas of ongoing research ranging from large scale sequencing projects to expression studies related to tissue type or environmental conditions or epigenomics. Data is then made available through several Entrez databases. NCBI's Entrez search system allows for searches between databases including the PubMed database. The presentation will include a summary of available data of different types and will include information on data access.

A subset of **epigenetics**-specific data from the GEO and SRA databases is processed and mapped to genomic coordinates to generate tracks that provide a visual representation of the data. Colgl_endopserm_meDIP-seq 🎎 Colgl_dme_endosperm_meDIP-seq misc feature Features Ler_embryo_meDIP-seq Col0_wholeplant_meDIP-seq Embryo_BS_seq WT_endosperm_BS_seq dme_endosperm_BS_seq

The photoreceptor chaperone FHY1 has an independent role in gene modulation and plant development under far-red light [ChIP-seq] (thale cress) To incorporate the far-red light (FR) signal into a strategy for optimizing plant growth, FAR-RED ELONGATED HYPOCOTYL1 (FHY1) mediates the nuclear translocation of the FR photoreceptor phytochrome A (phyA) and facilitates the association of phyA with the promoters of numerous associated genes crucial for the respo GSM1400504: FHY1-input; Arabidopsis thaliana; ChIP-Seq 1 ILLUMINA (Illumina HiSeq 2000) run: 208.4M spots, 15.8G bases, 10.3Gb downloads Accession: SRX555092 Attributes: Scope: Multiisolate; Material: Genome; Capture: Whole; Method type: Sequencing GSM1400503; FHY1-GFP 3 ChIP-seq; Arabidopsis thaliana; ChIP-Seq 1 ILLUMINA (Illumina HiSeq 2000) run: 203.8M spots, 15.5G bases, 10.1Gb downloads Other Accessions: GEO:GSE58083 Project Data: GSM1400502: FHY1-GFP 2 ChIP-seq; Arabidopsis thaliana; ChIP-Seq Number Resource Name of Links SEQUENCE DATA

1 ILLUMINA (Illumina HiSeg 2000) run: 202,6M spots, 15,4G bases, 10,1Gb downloads SRA Experim GSM1400501: FHY1-GFP 1 ChIP-seq; Arabidopsis thaliana; ChIP-Seq PUBLICATIONS 1 ILLUMINA (Illumina HiSeq 2000) run: 271.6M spots, 20.6G bases, 12.9Gb downloads PubMed OTHER DATASETS chaperone FHY1 FHY1-input GEO DataSets Arabidopsis thaliana Generic sample ▼ GEO Data Details BioSample: SAMN02803822: GEO: GSM1400504: SRA: SRS82238 BioProject SRA GEO DataSets Data volume, Supplementary Mbytes FHY1-GFP 3 ChIP-sec Arabidopsis thaliana Generic sample BioSample: SAMN02803816; GEO: GSM1400503; SRA: SRS62238 BioProject SRA GEO DataSets association of phyA with the promoters of numerous associated genes crucial for the response to environmental stimuli. Howeve FHY1-GFP 2 ChIP-seq Arabidopsis thaliana Generic sample BioSample: SAMN02803815: GEO: GSM1400502 Download data: GEO (BED), SRA SRP042366 BioProject SRA GEO DataSets

The **Gene** database serves as a central hub for gene-centric information. It allows for rapid access to Reference Sequences and graphical displays of genomic maps.

The RefSeqs for a Gene record are presented in an interactive and customizable graphical Sequence Viewer. Regions of interest can be explored using the "Tools" menu. Sequence can be viewed, downloaded or BLAST aligned. Configure button allows for various tracks to be aligned and viewed in a single window.

General protein information Tracks Custom Data \longrightarrow \parallel \parallel \parallel Click on track to display settings. To re-order track XM_002891044. Active Tracks Lineage Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Gunneridae; Pentapetalae; rosids; malvids; Brassicales; Brassicaleae; Camelineae; Arabidop 🔍 Search Tracks Also known as ATNHX8; F10B6.1; F10B6_1; Na+/H+ exchanger 8; NHX8; SODIUM HYDROGEN EXCHANGER 8 ✓ SNP Related information XM_009150599,1 misc feature Feature Variation BLAST Results for: ref[NC_003070.9] (30427671 letter NM 101333.4 Genes/Product XM_002892774.1 Features Uploaded Data some: 1: NC 003070.9 (5031950..5036477, complement Full text in PMC BLAST Full text in PMC_nucleotide Gene neighbor 5026364 Genomic regions, transcripts, and products Go to reference sequence details Primer BLAST allows for the selection of primer NC 003070.9: 5.0M..5.0M (5.9Kbp) C ▼ Find sequences within a selected region. RefSeq RNAs

BLAST and Primer Search

Download

Preferences

Printer-Friendly PDF

Genomic regions, transcripts, and produc

Links to other resource

BLAST Search (Visible Range)

Primer BLAST (Visible Range)

BLAST Search (Selection)

Primer BLAST (Selection)

General informatio

FAQ

Help

FTP site

BioProiec

Map View

Feedback

Contact Help De

Submit Correction

Submit GeneRIF

Subscription

Recent activity

NHX8 [Arabidopsis thaliana]

Q KN708641 : KN711235[PACC] (2595)

GEO

etailed primer reports Forward primer GCCAATGGCCAAACTCCTGA 5032178 5032197 60.90 55.00 7.00

Reverse primer

Product length

Primer pair

Reverse primer

Product length

RefSeq records are linked to the **UniGene** database that connects expressed sequences from GenBank and RefSeqs into gene clusters

5032265 5032285 59.99 52.38 4.00

5034251 5034232 59.97 55.00 6.00

based on sequence similarity. Nelumbo nucifera, whole genome shotg hale cress protein-coding gene NHX8. Represented by 6 ESTs from 3 cDNA libraries. Corresponds to 2 reference equences (different isoforms). [UniGene 12901 - At.27734] nelumbo nucifera[orgn] AND and wgs_n At.27734 0/6299 floral merister 0/4229 sodium/hydrogen exchanger 8 A. thaliana 0/62048 Other hits (2 of 10) [Show all] 0/12182 S. bicolor 0/49512 0/16148 0/13014 vegetative tissue esults: 1 to 20 of 90 organism: Arabidopsis thaliana Reporter: GPL483, 524988 (ID_REF), GDS481, AT1G14880 (ORF) organism: Arabidopsis thaliana Reporter: GPL483, 524988 (ID_REF), GDS480, AT1G14880 (ORF) Arabidopsis thaliana NHX8 protein (NHX8) mRNA, complete cds, alternatively spliced EST sequences (6

The Map Viewer is the primary graphical display tool for eukaryotic genomes. Both sequence and non-sequence maps can be displayed, aligned and searched.

a_cpa1; Na+/H+ antiporter, bacterial form

mRNA: NM 101333.4

Graphical View: NM 101333.

BLAST mRNA: NM 1013

Links & Tools

Publications describing functional annotation of a

Gene can be added via GeneRIF

BEST Arabidopsis thaliana protein match is: sodium proton exchanger, putative (NHX7) (SOS1) (TAIR:AT2G01980.1

The following sections contain reference sequences that belong to a specific genome build. Explai

Range 5031950..5038477

See proteins identical to NP 001184999

See proteins identical to NP 172918.3

UniProtKB/Swiss-Prot Q3YL57

Status: REVIEWED

Status: REVIEWED

Conserved Domains (2) summary

Download GenBank, FASTA, Sequence Viewer (Graphics)

NM 101333.4 -> NP 172918.2 sodium/hydrogen exchanger 8 [Arabidopsis thaliana

Title: mRNA-sodium/hydrogen exchanger (

BLAST Genomic: NC 003070.9 (5,031,972..5,036,448)

Na+/H+ exchanger 8 (NHX8); CONTAINS InterPro DOMAIN/s: Cation/H+ exchanger, conserved region (InterPro:IPR018422), Cation/H+ exchanger (InterPro:IPR006153), Na+/H+ exchanger, isoforms 7/8, conserved region (InterPro:IPR018422)

Na+/H+ exchanger 8 (NHX8); CONTAINS InterPro DOMAIN/s: Cation/H+ exchanger, conserved region (InterPro:IPR018422), Cation/H+ exchanger (InterPro:IPR008153), Na+/H+ exchanger, isoforms 7/8, conserved region (InterPro:IPR018418)

Reference Sequence

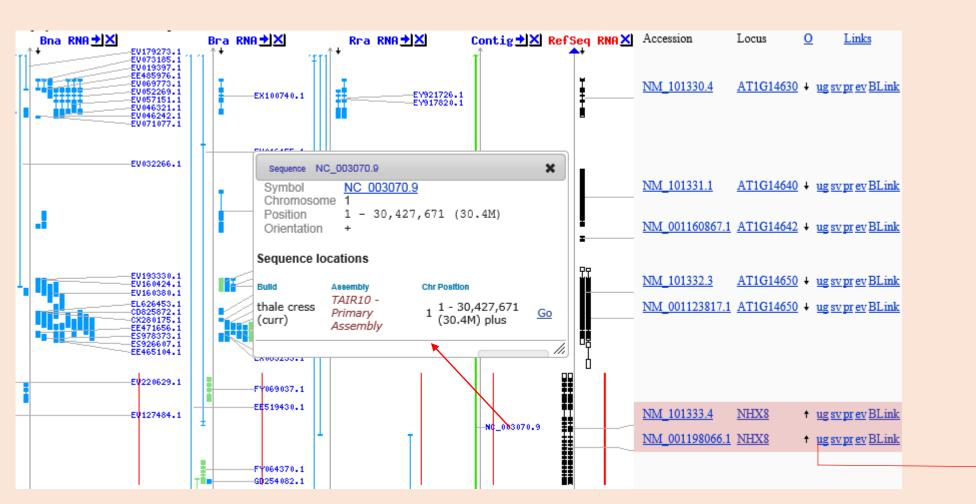
GenBank View: NC 003070.9 (5,031,972...5,036,448), NM 101333.4

FASTA View: NC 003070.9 (5,031,972..5,036,448), NM 101333.4

Location: complement(5,031,972..5,036,448)

laster Map: Genes On Sequence egion Displayed: 4,882K-5,186K bp At UniG**⇒**|X MIPS TAIR SIGNAL sv pr d1 hm external 60S ribosomal protein L10-1 AT1G14330 * MIPS TAIR SIGnAL sv prd1hm external F-box/kelch-repeat protein SKIP11 + MIPS TAIR SIGNAL svprdlhm external FOUR LIPS transcription factor R2R3-MYB MIPS TAIR SIGNAL svprd1hm external UDP-GALACTOSE TRANSPORTER 3 APK2A MIPS TAIR SIGNAL sv pr dl hm external PROTEIN KINASE 2A UBC1 * MIPS TAIR SIGNAL sv pr dl hm external ubiquitin-conjugating enzyme E2 1 * MIPS TAIR SIGNAL svprdlhm external ssDNA-binding transcriptional regulator MIPS TAIR SIGNAL sv pr d1 hm external putative pectate lyase 3 MIPS TAIR SIGNAL sv pr dl external NADH dehydrogenase (ubiquinone) MIPS TAIR SIGnAL sv pr dl hm external protein alfin-like 7 * MIPS TAIR SIGNAL sv pr dl hm external inositol oxygenase 1 * MIPS TAIR SIGNAL sv pr d1 hm external hypothetical protein AT1G14560 + MIPS TAIR SIGnAL syprdlhm external Mitochondrial substrate carrier family protein AT1G14570 + MIPS TAIR SIGNAL sv pr d1 hm external UBX domain-containing protein * MIPS TAIR SIGNAL syprdlhm external Valyl-tRNA synthetase MIPS TAIR SIGNAL sv pr dl hm external protein decoy MIPS TAIR SIGnAL svprdl external sodium/hydrogen exchanger 8

The "Maps and Options" dialog allows for the display of multiple maps. Genome assembly information can be found by selecting the Contig map. The image below displays the A. thaliana, B. napus, B. rapa and R. raphanistrum transcript maps aligned to A. thaliana genome.

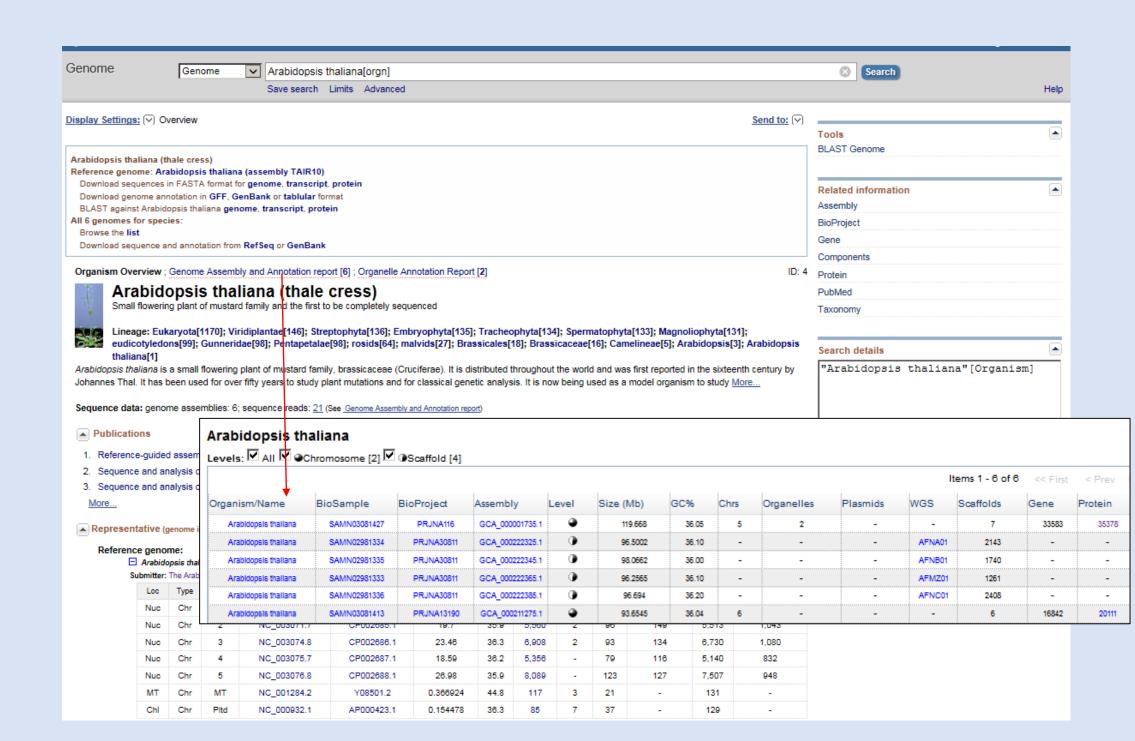


The **BioProjects** database describes large-scale research efforts, ranging from genome and transcriptome sequencing projects to epigenomic and variation analyses. It provides an organizational framework to access data across multiple resources and multiple submission time points. As of now, close to 2,300 BioProjects are registered for Arabidopsis of which 2,234 are for A. thaliana.

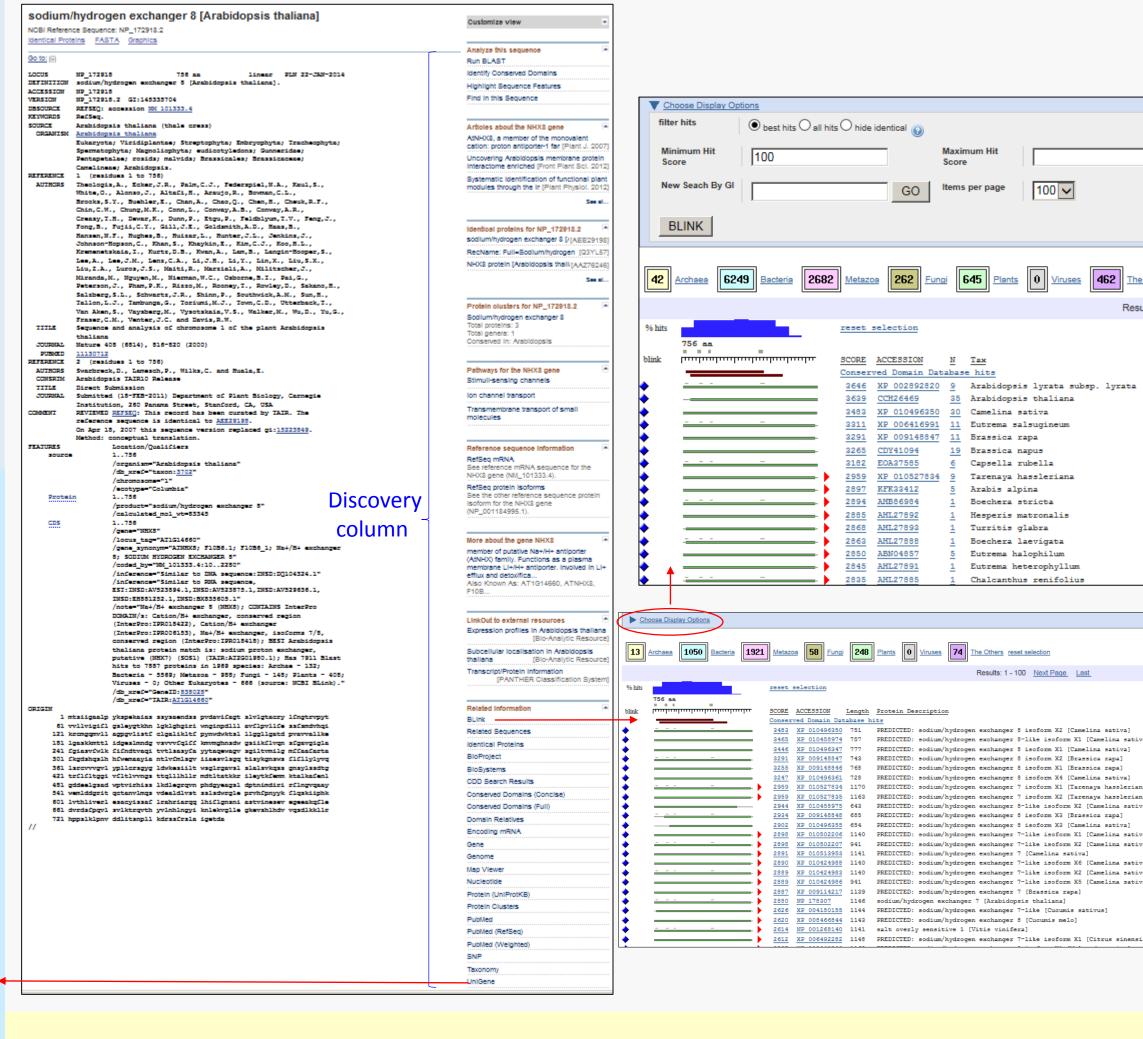
The **Epigenomics** database is a comprehensive resource for whole-genome epigenetic data sets. Epigenomics BioProjects are linked to the raw data in the SRA and GEO databases.

✓ Arabidopsis[orgn] Arabidopsis thaliana strain:Col-0 Arabidopsis thaliana strain:Col-0 Epigenomic Show additional filters Display Settings: 🗸 Summary, 20 per page, Sorted by Default order Organism: Arabidopsis thaliana Taxonomy: Arabidopsis tha Project data type: Epigenom Project Types See also 3 genomes matching your organism search Attributes : Scope: Multiisok Primary submission (2264) University of Oxford Epigenomics BioProjects RefSeq (2) Results: 1 to 20 of 2264 Data Types Transcriptional profiling of the Arabidopsis root quiescent center plant development under far-red light [ChIP-seq Clone ends (1) Organism: Arabidopsis thaliana Organism: Arabidopsis thaliana Epigenomics (189) Taxonomy: Arabidopsis thaliana (thale cress) Taxonomy: Arabidopsis thaliana (thale cre Project data type: Transcriptome or Gene expression Genome sequencing (44) Attributes: Scope: Multiisolate; Material: Transcriptome; Capture: Whole; Method Type: Arra Attributes: Scope: Multiisolate; Material: Genome; Capture: Whole; Method Type: Sequencing Map (3) MCDB, Yale University Other (80) Accession: PRJNA248878 ID: 248878 Accession: PRJNA270538 ID: 270538 Phenotype/genotype (1) Targeted Jocus (6) Arabidopsis thaliana strain:Colombia-0 Organism: Arabidopsis thaliana Transcriptome (1915) Arabidopsis thaliana strain:Colombia-0 Targeted Locus (Loci) Taxonomy: Arabidopsis thaliana (thale cress Variation (23) Organism: Arabidopsis thaliana Project data type: Epigenomics Taxonomy: Arabidopsis thaliana (thale cress) Project data type: Targeted Locus (Loci) Project Data Arabidopsis lyrata Attributes : Scope: Multiisolate; Material: Genome; Capture: Targeted Locus/Loci; Method Type Nucleotide (9) University of Washington Arabidopsis lyrata is an out-crossing diploid, closely related to A. thaliana. Protein (6) Accession: PRJNA270258 ID: 270258 Kingdom: Eukaryota; Subgroup: Land Plants Sequence data: genome assemblies:2 Assembly (8) Date: 2009/11/30 Arabidopsis thaliana SRA (376) ID: 493 RNAseq of PhyAB mutants for Red light GEO DataSets (2109) Taxonomy: Arabidopsis thaliana (thale cress) Project data type: Other **Organism Groups** Attributes: Scope: Other; Material: Other; Capture: Other; Method Type: Sequencing Model plant for studying metal accumulation Plants (2264) Hanada lab Kingdom: Eukaryota; Subgroup: Land Plants Accession: PRJDB2760 ID: 269344 Sequence data: genome assemblies: Date: 2014/01/16 Clear all ID: 13155 Arabidopsis thaliana Small flowering plant of mustard family and the first to be completely sequenced Kingdom: Eukaryota; Subgroup: Land Plants Sequence data: genome assemblies:6

> The **Genome** database provides organism specific genomic information with links to assembly and annotation information.



Reference Sequence (RefSeq) protein records are in the Protein database. The Discovery column on each RefSeq record provides access to a wealth of information from related databases. Homologs can be identified following the "Blink" in the protein record. "Choose Display Options" provides filters to show single or unique protein from each organism.



The **Taxonomy** database not only provides phylogenetic lineages of more than 160,000 organisms that have data in the various NCBI databases, but also provides a quick overview and access to the data for a particular organism. The nucleotide sequences for A. thaliana are present in three of the databases: Nucleotide, EST, and GSS. Reference Sequence RNA and genomic records are in the Nucleotide database.

